



United States Patent & Trademark Office; U.S. DEPARTMENT OF COMMERCE

PTO/SB/33 (07-05)

<b>APPEAL BRIEF REQUEST FOR REVIEW</b>		Docket Number (Optional) 058268.00224
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to "Mail Stop AF, Commissioner of Patents, P.O. Box 1450, Alexandria, VA 22313-1450" [37 CFR 1.8(a)]  on _____  Signature _____  Typed or printed Name _____	Application Number:  10/660,780	Filed: September 12, 2003
	First Named Inventor:  Nambi SESHADRI	
	Art Unit: 2626	
	Examiner: Lerner, Martin	

**Mail Stop AF**  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Applicant requests review of the final rejection in the above-identified application. No amendments are being filed with this request.

This request is being filed with a Notice of Appeal.

The review is requested for the reason(s) stated on the attached sheet(s).

Note: No more than five (5) pages may be provided.

I am the

- ☐ Applicant/Inventor.  
☐ assignee of record of the entire interest.

See 37 CFR 3.71. Statement under  
37 CFR 3.73(b) is enclosed

- ☒ Attorney or agent of record.  
Registration No. 51,091

- ☐ Attorney or agent acting under 37 CFR 1.34.  
Reg. No. is acting under 37 CFR 1.34 \_\_\_\_\_

Signature

David E. Brown

Typed or printed name

(703) 720-7883

Telephone number

December 21, 2007

Date

NOTE: Signatures of all of the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below\*.

- ☒ \*Total of 1 forms are submitted.



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

Confirmation No.: 5880

Nambi SESHADRI

Art Unit: 2626

Application No.: 10/660,780

Examiner: Martin Lerner

Filed: September 12, 2003

Attorney Dkt. No.: 058268.00224

For: CORRELATING VIDEO IMAGES OF LIP MOVEMENTS WITH AUDIO  
SIGNALS TO IMPROVE SPEECH RECOGNITION

**PRE-APPEAL BRIEF REQUEST FOR REVIEW**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

December 21, 2007

Sir:

In accordance with the Pre-Appeal Brief Conference Pilot Program guidelines set forth in the July 12, 2005 Official Gazette Notice, Applicant hereby submits this Pre-Appeal Brief Request for Review of the final rejections of claims 1-21 in the above identified application. Claims 1-21 were finally rejected in the Office Action dated September 21, 2007. Applicant filed a Response to the Final Office Action on November 8, 2007, and the Office issued an Advisory Action dated November 23, 2007 maintaining the final rejections of claims 1-21. Applicant hereby appeals these rejections and submits this Pre-Appeal Brief Request for Review.

The Office Action rejected claims 1-3, 5-7, 9-11, and 13-15 under 35 U.S.C. 103(a) over US Patent No. 6,526,395 to Morris (Morris), in view of US Patent No 6,931,351 to Verma et al. (Verma). Applicants submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the pending claims. This failure constitutes clear error in the Office Action.

Claim 1, from which claims 2-4, 16, and 19 depend, is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. It is determined if the audio signals can be processed. Based on the detection that at least a portion of audio signals can not be processed, the video signals are processed. At least one of the audio signals and the video signals are converted into recognizable information. A task is implemented based on the recognizable information.

Claim 5, from which claims 6-8, 17, and 20 depend, is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from a speech source. A video signal receiver is configured to receive video signals from the speech source. A processing unit is configured to detect if the audio signals can be processed and if so, to process the audio signals. The video signals are processed based on the detection that at least a portion of the audio signals cannot be processed. A conversion unit is configured to convert at least one of the audio signals and the video signals to recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 9, from which claims 10-12, 18, and 21 depend, is directed to a system for speech recognition. A first receiving means is configured for receiving audio signals from a speech source. A second receiving means is configured for receiving video signals from the speech source. A processing means is configured for detecting if the audio signals can be processed and processing the audio signals if the audio signals can be processed. The processing means processes the video signals based on the detection that at least a portion of the audio signals can not be processed. A converting means is configured for converting at least one of the audio signals and the video signals to recognizable information. An implementing means is configured for implementing a task based on the recognizable information.

Claim 13 is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. If the

audio signals can be converted into a recognizable format, the audio signals are processed. The audio signals are converted into recognizable information. The video signals are processed when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. The processed video signals are converted into the recognizable information. A task is implemented based on the recognizable information.

Claim 14 is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from a speech source. A video signal receiver is configured to receive video signals from the speech source. A first processing unit is configured to detect if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed. A first conversion unit is configured to convert the audio signals to recognizable information. A second processing unit is configured to process the video signals when the audio signals cannot be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second conversion unit is configured to convert the processed video signals into the recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 15 is directed to a system for speech recognition. A first receiving means receives audio signals from a speech source. A second receiving means receives video signals from the speech source. A first processing means detects if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed. A first converting means converts the audio signals into recognizable information. A second processing means processes the video signals when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second converting means converts the processed video signals into the

recognizable information. An implementing means implements a task based on the recognizable information.

Applicants submit that each of the pending claims recites features that are neither disclosed nor suggested in the cited references.

As discussed in Applicants' previous correspondence, Morris is directed to an apparatus includes a video input unit and an audio input unit. The apparatus also includes a multi-sensor fusion/recognition unit coupled to the video input unit and the audio input unit, and a processor coupled to the multi-sensor fusion/recognition unit. The Office Action admits that Morris failed to disclose the feature of "detecting if the audio signal can be processed," "processing the audio signals if it is detected that the audio signals can be processed," and "processing the video signals if it is detected that at least a portion of the audio signal cannot be processed." The Office Action relied on Verma to cure these deficiencies.

Verma is directed to decision making in classification problems. Verma describes classifying samples to one of a number of predetermined classes using a number of class models or classifiers to form order statistic for each classifier. Verma describes that audio and video vectors are similarly processed. The weight for the audio is determined and since there are only two classifiers, the weight for video is determined as a compliment of the weight for the audio as the linear summation of all weights is "1". The threshold is defined for sample confidence values of audio. The class confidence value if the audio is checked against its threshold. If this test is passed, the audio weight is computed as a constant term and a term which is dependent on the overall confidence of the audio channel. If the test is failed, the constant term changes. See col. 4 lines 41-51 of Verma.

Applicants respectfully submit that the cited references fail to disclose or suggest at least the feature of "processing the video signals based on a detection that at least a portion of the audio signal cannot be processed," as recited in claims 1, 5, and 9. Specifically, Applicants respectfully submit that Verma fails to cure the admitted deficiencies of Morris. This failure constitutes clear error in the Office Action.

Verma fails to disclose or suggest that the processing of the video vector is based on the inability to process a portion of the audio vector. As discussed in Applicant's previous correspondence, Verma merely describes that the video vector is processed in a similar way to the audio vector, and the weights of the audio and video signal are complimentary.

Still further, Verma does not disclose or suggest that there is a determination whether at least a portion the audio signal can be processed and if not, processing the video signal. For example, Verma does not disclose that if the confidence level of the audio signal is "0" then the video signal is processed instead of the audio signal. This is further evidenced in Fig. 2 of Verma, which merely illustrates that the assigning of weights to the audio 110 and video signals 120.

Thus, Applicants respectfully submit that Verma fails to cure the admitted deficiencies of Morris. Thus, the cited references fail to disclose or suggest all of the features recited in claims 1, 5, and 9. This failure constitutes clear error in the Office Action.

Regarding claims 13-15, Applicants respectfully submit that the cited references fail to disclose or suggest at least the feature of "wherein the video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information." In other words, as previously discussed, Verma is silent with regards to processing the video signals that coincide with the portion of the audio signal that can not be converted. As discussed above, Verma merely describes assigning weights to the audio and video signals, and no determination is made as to whether a portion of the audio signal can be converted before proceeding to processes the associated video signal. Thus, in regards to claims 13-15, Verma fails to cure the admitted deficiencies of Morris. This failure constitutes clear error in the Office Action.


The Office Action rejected claims 4, 8, and 12 under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in view of Bakis. Claims 16-18 were rejected under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in further view of Basu. Claims

19-21 are rejected under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in further view of Brunelli. For the reasons discussed in previous correspondence, each of Bakis, Basu and Brunelli fails to cure the significant deficiencies of Morris and Verma.

Applicants submit that the Office Action failed to establish *prima facie* obviousness in rejecting each of claims 1-21. This failure constitutes clear error in the Office Action.

Reconsideration and withdrawal of the rejections, in view of the clear errors in the Office Action, is respectfully requested. In the event this paper is not being timely filed, the applicant respectfully petition for an appropriate extension of time. Any fees for such an extension together with any additional fees may be charged to Counsel's Deposit Account 50-2222.

Respectfully submitted,

  
\_\_\_\_\_  
David E. Brown  
Registration No. 51,091

**Customer No. 32294**  
SQUIRE, SANDERS & DEMPSEY LLP  
14<sup>TH</sup> Floor  
8000 Towers Crescent Drive  
Tysons Corner, Virginia 22182-2700  
Telephone: 703-720-7800; Fax: 703-720-7802

DEB/jkm:cqc

Enclosures: PTO/SB/33 Form  
Notice of Appeal  
Check No. 017763